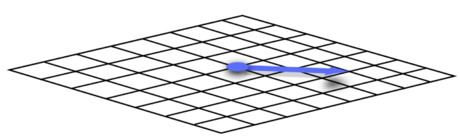
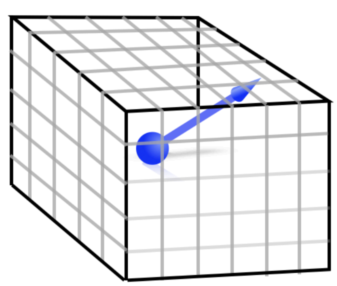
The Core Of Linear Algebra

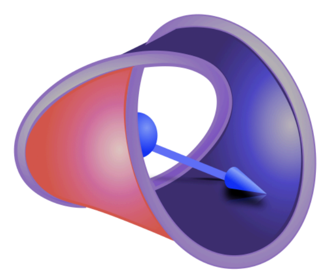
by DiBeos

Table of Contents

Vectors can exist in $1D$, in $2D$, in $3D$, or in any higher dimensions:







Of course, these vectors live in a space, which is usually called a *vector space*. Vectors are not necessarily arrows, though. It is just that arrows are their most classic representation. Vectors are actually any mathematical object that satisfies a list of conditions defined by its vector space $\boldsymbol{V}$:

($\forall\, u, v, w \in \boldsymbol{V} \quad \wedge \quad \forall\, a, b \in \mathbb{F}$ (field))

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$$
$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$
$$\exists \mathbf{0} \in \mathbf{V} : \quad \mathbf{u} + \mathbf{0} = \mathbf{u}$$
$$\forall \mathbf{u} \in \mathbf{V}, \exists - \mathbf{u} \in \mathbf{V} : \quad \mathbf{u} + (-\mathbf{u}) = \mathbf{0}$$
$$\forall a \in \mathbb{F} \text{ (field)} \wedge \forall \mathbf{u} \in \mathbf{V}, \; a\mathbf{u} \in \mathbf{V}$$
$$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$$
$$a(b\mathbf{u}) = (ab)\mathbf{u}$$
$$(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$$
$$\exists 1 \in \mathbb{F} : \quad 1\mathbf{u} = \mathbf{u}$$

Examples of objects that satisfy these conditions, and therefore are qualified to be called vectors, are:

- **Matrices**

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \; \begin{bmatrix} 0 & -1 \\ 2 & 3 \end{bmatrix} \; \begin{bmatrix} 4 & 3 & 2 \\ 7 & -10 & 2 \\ 2 & 1 & 0 \end{bmatrix}$$

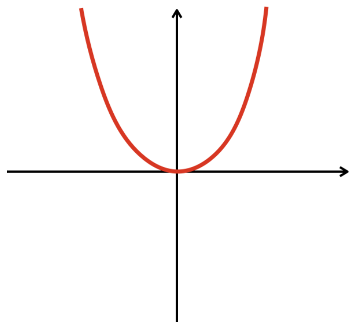
- **Polynomials**

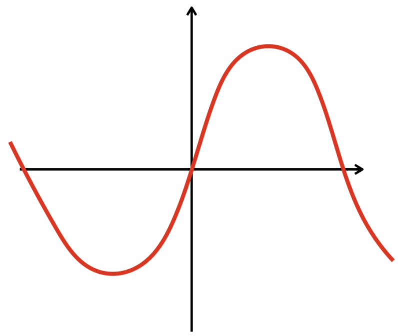
$$p(x) = 2x^7 + 4x^6 + \ldots + x + 2$$

$$q(x) = x^2 - 4$$

- **Functions**

$$f(x) = x^2 \qquad g(x) = \sin(x)$$





- **Sequences**

$$(a)_{n \in \mathbb{N}} = (a_1, a_2, \ldots)$$

$$1, 1, 2, 3, 5, 8, 13, 21, 34$$

- **Complex numbers**

$$z = a + ib;\ \ a, b \in \mathbb{R}$$

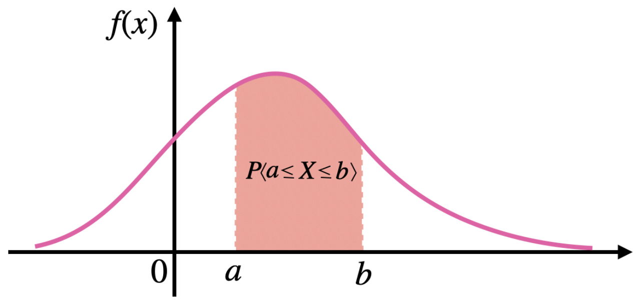
$$i = \sqrt{-1}$$

- **Solutions to linear differential equations**

$$y^{(n)} + 2y^{(n-1)} + y' = 0$$

- **Probability distributions**

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



More specifically, probability densities.

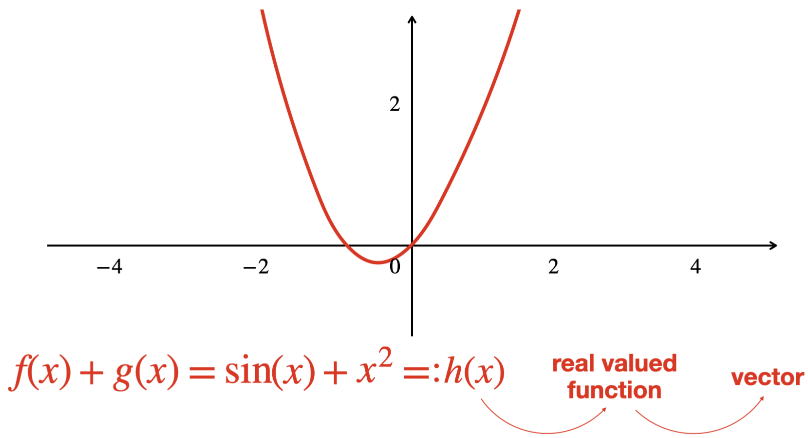
The core characteristics of a vector are:

1. Vector Addition
2. Scalar Multiplication

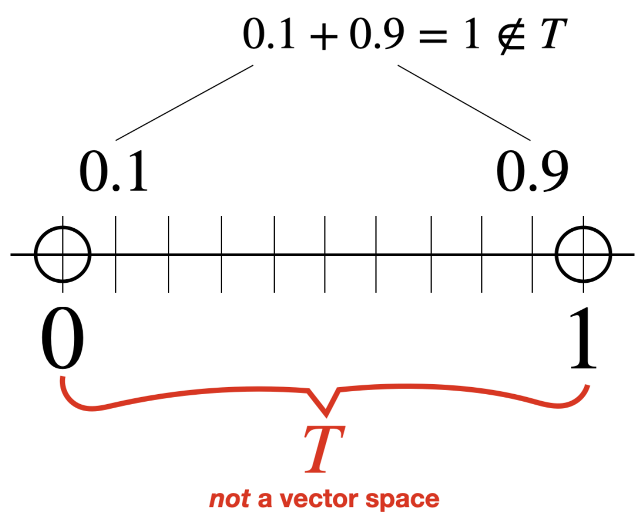
Let us see a few quick examples:

The set of all *real-valued functions* $f: \mathbb{R} \rightarrow \mathbb{R}$ is a vector space, with its elements (in this case, functions) identified as vectors.

Pick 2 such functions $f(x) = sin(x)$ and $g(x) = x^2$. The addition $f(x) + g(x) = sin(x) + x^2 =: h(x)$ creates a new function in the same vector space, because the resultant function $h(x)$ is still a real-valued function, and therefore it belongs to the original space.



As a counterexample, we notice that the set of all real numbers between 0 and 1, $T = \{x \in \mathbb{R} \mid 0 < x < 1\}$, is <u>not</u> a vector space, because even though 0.9 and 0.1 belong to T, the sum $0.9 + 0.1 = 1 \notin T$. We say that T is <u>not</u> closed under addition.



The set of all real-valued functions $f: \mathbb{R} \to \mathbb{R}$ also satisfies scalar multiplication, i.e. it is closed under scaling. So, if we pick a real number $a \in \mathbb{R}$, then multiplying a (say $a = 2$) by $f(x) = sin(x)$ results in the function $r(x) := 2 \cdot sin(x)$ which is a real-valued function. Beyond that, multiplying a by the sum $f(x) + g(x)$ is the same as multiplying a by $f(x)$ and then adding it with the multiplication of a times $g(x)$:

$$a \cdot (f(x) + g(x)) = a \cdot f(x) + a \cdot g(x)$$

$$2 \cdot \left(sin(x) + x^2\right) = 2 \cdot sin(x) + 2 \cdot x^2$$

This property is called *distributivity*.

Furthermore, if we pick yet another real number $b \in \mathbb{R}$, we notice that scalars are *compatible* with real-valued functions. For example, for $b = \frac{1}{3}$:

$$(a \cdot b) \cdot f(x) = a \cdot (b \cdot f(x))$$

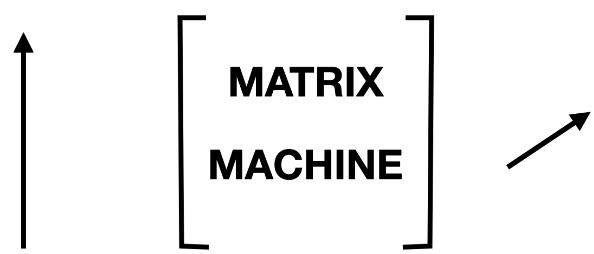
$$\left(2 \cdot \frac{1}{3}\right) sin(x) = 2 \cdot \left(\frac{1}{3} \cdot sin(x)\right)$$

Finally, we notice that there is a unique *identity* function ($I(x) = 1$) such that $I(x) \cdot f(x) = f(x)\ ,\ \forall f(x)$.

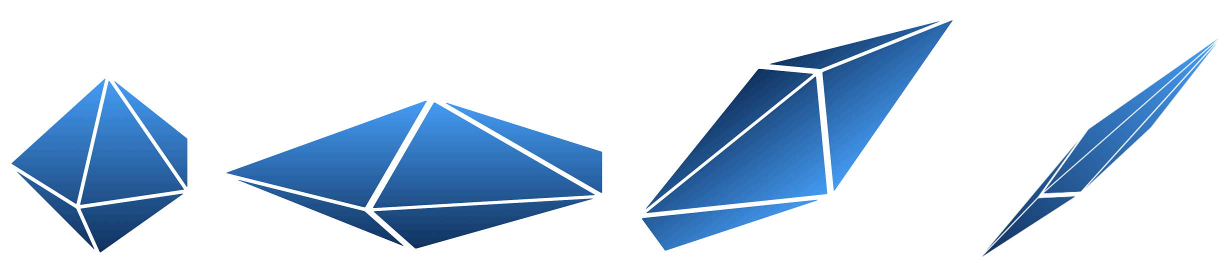
Scalar multiplication seems obvious, but not all sets and spaces satisfy it. As a counterexample, the set of integers modulo n $\left(i.e.\ \mathbb{Z}_n = \{0, 1, \dots, n - 1\}\right)$ fails scalar multiplication, and therefore its elements cannot be called vectors. Pick $\mathbb{Z}_5 = \{0, 1, 2, 3, 4\}$, for instance, and the real number 0.5 as our scalar. If we perform the operation 0.5 times $3 \in \mathbb{Z}_5$, the result $\frac{3}{2} \notin \mathbb{Z}_5$. We say that $\mathbb{Z}_5$ is <u>not</u> closed under scaling.

We will talk more about vectors, especially how to *transform* them and the concept of *linear independence*, but first we would like to give you a general picture of what *Linear Algebra* is all about.

Linear Algebra is the study of how things can be measured, combined and transformed in spaces that have a specific structure, and at the same time preserve linearity – which means that many relations are *proportional* and *additive*. In practice, it deals with vectors (which often represent quantities like directions and magnitudes), and matrices (which act as "machines" that transform these vectors).

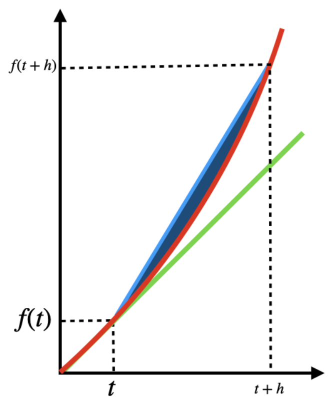


These transformations can involve stretchings, rotations, compressions and so on. But always in a linear way.



Linear Algebra gives us a framework to understand relations, like how systems of equations can interact, how data can be represented, and how changes in one part of the system affects the whole thing.

It is also a language that helps us to navigate higher dimensions, and model physical forces in nature, or even create very powerful computer algorithms. It is important to emphasize, though, that this subject is <u>only</u> concerned with transformations that preserve straight lines and proportionality. So, if your goal is to study more complex situations, like curves, sudden jumps, or exponential growth, then Linear Algebra is <u>not</u> the appropriate tool, and instead you would be in the realm of *non-linear mathematics*, for example: *calculus, differential equations, chaos theory, and so on…*



Calculus

$$\frac{dy}{dx} = f(x)$$

$$\frac{dy}{dx} = f(x, y)$$

$$x_1\frac{\partial y}{\partial x_1} + x_2\frac{\partial y}{\partial x_2} = y$$

Differential Equations

Chaos Theory

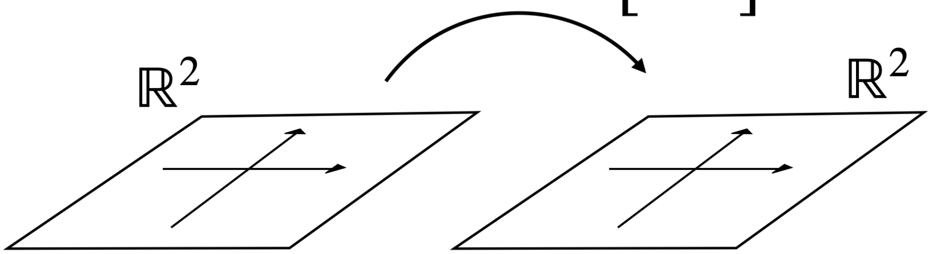
(You might wanna check out the other videos in the ***DiBeos*** *channel related to these subjects 😎)*

- Calculus is Impossible Without These 9 Things
- The 7 Indeterminate Forms that Changed Math Forever
- The Core of Dynamical Systems

Ok, now that you have a good intuition of what Linear Algebra aims to describe, let us get right into <u>Vector Transformations</u>:

Consider a transformation T, which is actually a mapping from the real plane to itself:

$$T : \mathbb{R}^2 \to \mathbb{R}^2 \qquad T = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$



This transformation can be represented as a matrix. Let us see why. Our goal is to transform the vector $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. In order to do it, we will *act* T on $\vec{x}$:

$$T\vec{x} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The result of multiplying a matrix by a vector is obtained by multiplying the first row of T to the vector $\vec{x}$, component by component, and then summing them up:

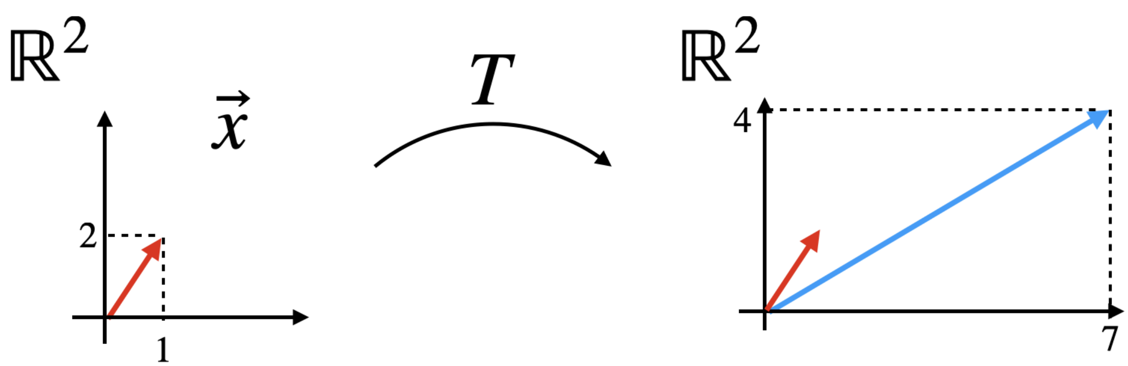
$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 + 1 \cdot 2 \\ \end{bmatrix}$$

Then we do the same with the second row:

$$\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 + 1 \cdot 2 \\ 1 \cdot 1 + 3 \cdot 2 \end{bmatrix}$$

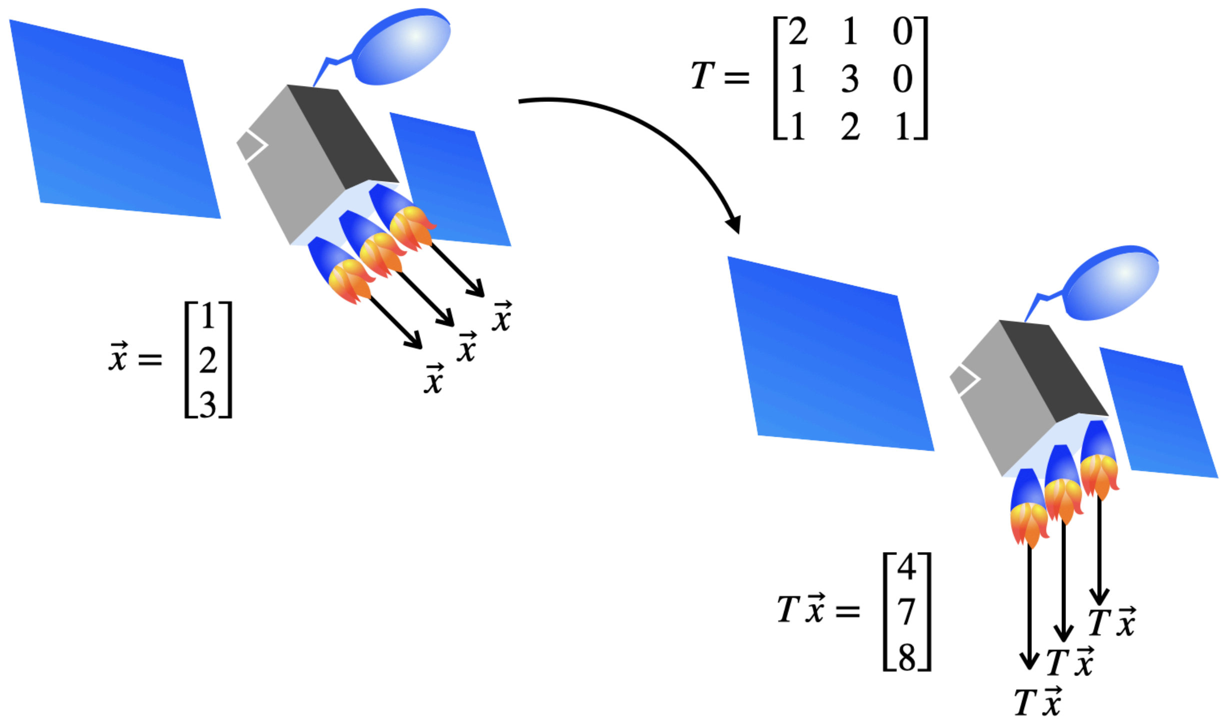
The result is the vector $T\vec{x}$:

$$T\vec{x} = \begin{bmatrix} 4 \\ 7 \end{bmatrix}$$



This is a very simple $2D$ transformation, but can be incredibly useful! Let us quickly illustrate it with a real world scenario.

Consider $\vec{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ as the force vector generated by the engine thrust of a satellite in its default direction and magnitude. The transformation $T = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 0 \\ 1 & 2 & 1 \end{bmatrix}$ can be a recalibration of the engine thrust system, for example. And $T\vec{x}$ would be the final force after fixing the engine misalignment.



As said earlier, a vector is not necessarily an arrow. It can be a polynomial, for example.

Consider the vector space of polynomials of degree at most 2:

$$P_2(\mathbb{R}) = \{\, p(x) = a_0 + a_1 x + a_2 x^2 \mid a_0, a_1, a_2 \in \mathbb{R} \,\}$$

This space is 3-dimensional, with basis $\{1,\ x,\ x^2\}$.

(We will talk more about basis later on, don't worry 😎)

Any polynomial $p(x) \in P_2(\mathbb{R})$ can be uniquely represented by a vector of its coefficients:

$$p(x) = a_0 + a_1 x + a_2 x^2 \quad \Longleftrightarrow \quad \vec{p} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

Let us define a linear transformation $T: P_2(\mathbb{R}) \rightarrow P_2(\mathbb{R})$ by specifying its action on the coefficients of $p(x)$:

$$T\vec{p} = \vec{q}$$

Since the space is $3D$, the transformation T is a 3×3 matrix (i.e. 3 rows and 3 columns), and $\vec{q}$ is the vector/polynomial that results from the transformation.

Ok, it is time to see a concrete example:

$p(x) = 3 + 2x - x^2$, which corresponds to the vector:

$$\vec{p} = \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix}$$

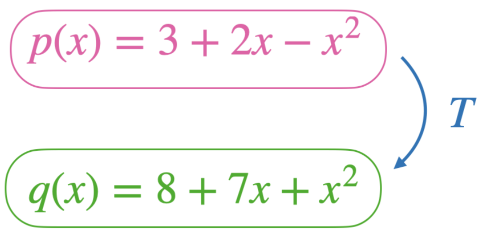
The transformation T is:

$$T = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & -1 \\ 1 & 0 & 2 \end{bmatrix}$$

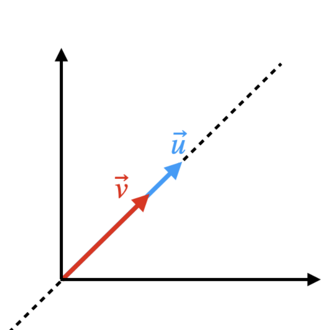
And after acting it on $\vec{p}$, we get the new vector $\vec{q}$.

$$T\vec{p} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 3 & -1 \\ 1 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ 1 \end{bmatrix} = \vec{q}$$

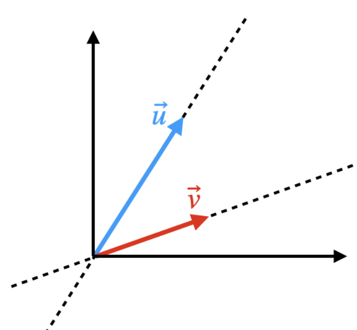
This vector corresponds to the polynomial $q(x) = 8 + 7x + x^2$.



Next, we talk about <u>Linear Independence</u>. Instead of going directly to its definition, let us stop for a moment and think about the semantics of these 2 words: *linear independence*, i.e. *not dependent* in a *linear way*. In particular, 2 vectors are linearly independent from each other if they do not lie on the same line through the origin. More precisely, when generalizing it to any set of vectors (which can be functions, polynomials, matrices, etc), we say that these vectors are linearly independent if none of them can be written as a *linear combination* of the others.



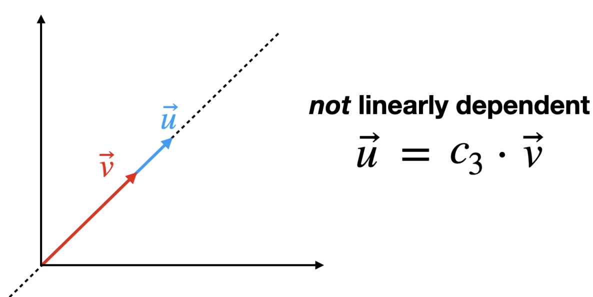
Linearly <u>Dependent</u>



Linearly <u>Independent</u>

In order to understand what a *linear combination* is, let us look for solutions to the equation $c_1\vec{u} + c_2\vec{v} = \vec{0}$, where c_1 and c_2 are scalars (i.e. 2 real numbers), and $\vec{u}$ and $\vec{v}$ are vectors. Let us also suppose that $\vec{u}$ and $\vec{v}$ are not null-vectors $\vec{0}$ (only zero entries).

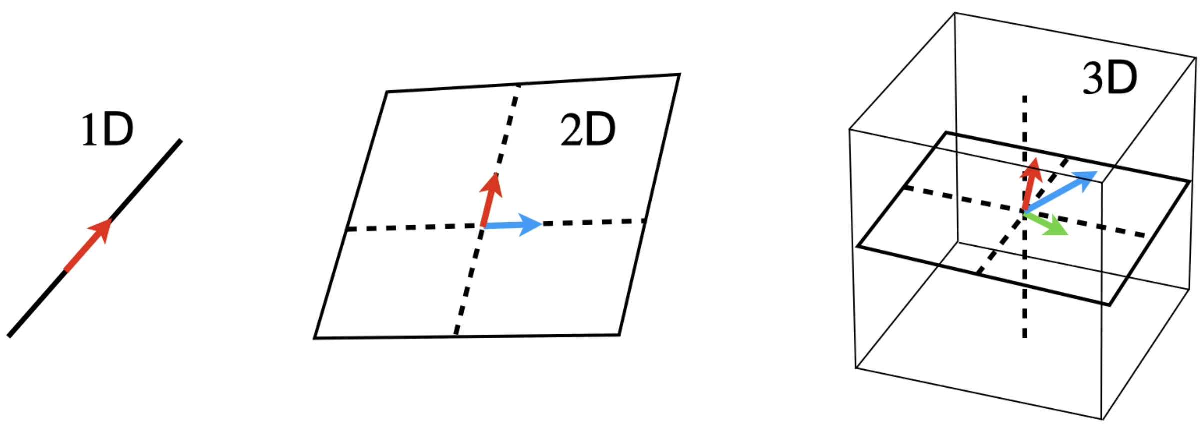
Now, if you try to find solutions c_1 and c_2 for this equation, and all you get is $c_1 = 0$ and $c_2 = 0$ (which are called *trivial solutions*), then it means that the 2 vectors, $\vec{u}$ and $\vec{v}$, are called linearly independent. Notice, though, that if there are non-trivial solutions (so, $c_1 \neq 0$ and $c_2 \neq 0$), then we can rewrite the equation as $\vec{u} = -\frac{c_2}{c_1}\vec{v}$. The new scalar $\left(-\frac{c_2}{c_1}\right)$ is still a real number, so we can represent it as $c_3 = -\frac{c_2}{c_1}$. What we just did was write the vector $\vec{u}$ as a rescaling of the vector $\vec{v}$: $\vec{u} = c_3 \cdot \vec{v}$, and therefore $\vec{u}$ and $\vec{v}$ must lie on the same line.



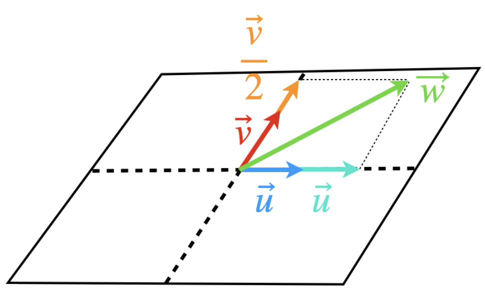
As a consequence, $\vec{u}$ and $\vec{v}$ are not linearly independent, and another equivalent way of expressing it is by saying that $\vec{u}$ can be written as a linear combination (i.e. a combination of linear terms) of $\vec{v}$, and vice-versa.

$$\vec{u} = c_3 \cdot \vec{v} \iff \vec{v} = \left(\frac{1}{c_3}\right) \cdot \vec{u}$$

The concept of linear independence is important because it allows us to build bases for vector spaces. A *basis* is a minimal set of vectors that can span the entire space around them.



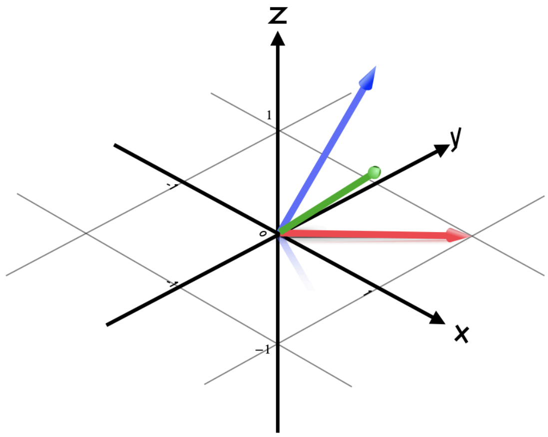
This means that every element in the space can be uniquely represented as a linear combination of the basis vectors.



$$\vec{w} = 2\vec{u} + \frac{3}{2}\vec{v}$$

Let us see some concrete examples:

Vectors in $\mathbb{R}^3$.



$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

These 3 vectors do form a basis for the vector space $\mathbb{R}^3$ because they are linearly independent of each other. We can check that.

Let $c_1, c_2, c_3 \in \mathbb{R}$ be scalars, then $c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 = \vec{0}$ implies that:

$$c_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + c_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

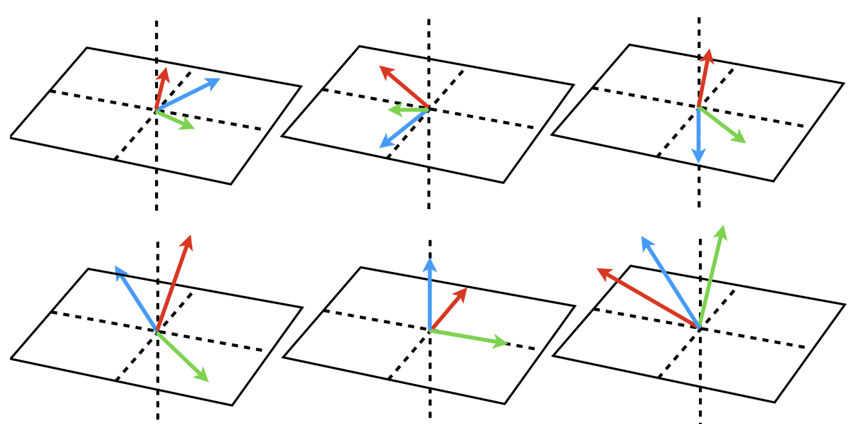
$$\Longrightarrow \begin{bmatrix} c_1 & + & c_3 \\ c_1 & + & c_2 \\ c_2 & + & c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Longrightarrow \begin{cases} c_1 + c_3 = 0 \\ c_1 + c_2 = 0 \\ c_2 + c_3 = 0 \end{cases}$$

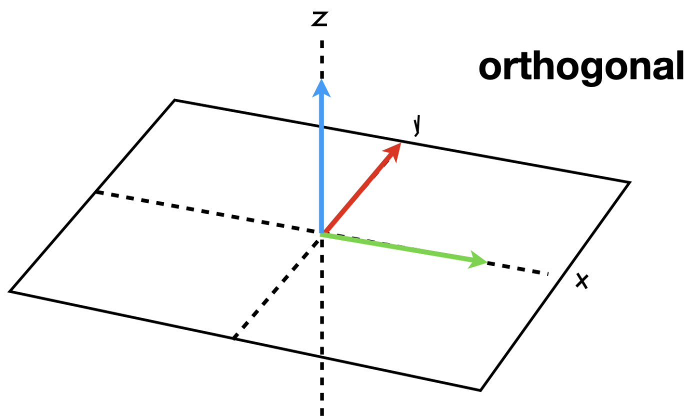
It turns out that the only solutions for this system of equations is the trivial one, in which $c_1 = c_2 = c_3 = 0$.

The conclusion is that $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are indeed linearly independent, and thus they qualify to be a basis of $\mathbb{R}^3$, i.e. any vector that you can think of in $\mathbb{R}^3$ can be decomposed into a linear combination of the "building blocks" of $\mathbb{R}^3$, which in this case are: $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

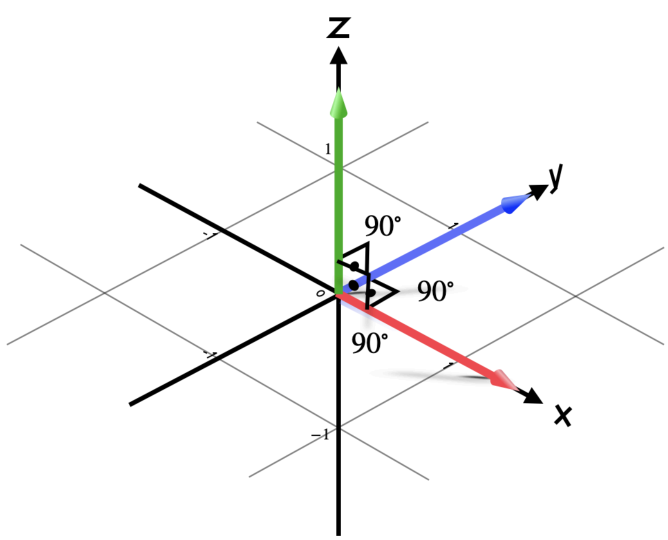
Of course, there are infinite choices of basis, but in order to qualify to be one of them, there must be at least 3 vectors involved, which comes from the fact that the dimension of the space $(\mathbb{R}^3)$ that we are spanning is 3.



A more natural choice of basis, though, would be a set of vectors that are not only linearly independent but also *orthogonal* to each other.



Again, they do not need to be orthogonal in order to constitute a basis, but it is a nice feature to have. First of all, what does it mean for 2 vectors to be orthogonal to each other? Intuitively (at least when using the arrow representation of a vector), 2 vectors are orthogonal to each other if the angle formed between them is 90°.



The problem with this intuition is that it fails when we treat vectors more generally (in a more abstract way), because vectors can be matrices, functions, complex numbers, polynomials, etc, and it doesn't make much sense to talk about the angle formed between 2 functions, for example.

So, in order to properly define orthogonality, first we need to introduce one of the most important operations that a vector space can have, namely the *dot product*.

Pick 2 vectors $\vec{u}$ and $\vec{v}$, and perform the dot product of them:

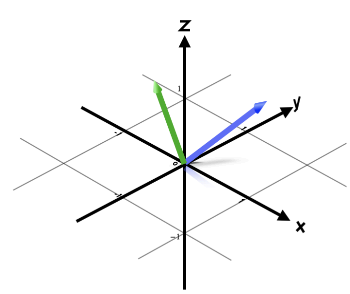
$$\vec{u} \cdot \vec{v} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

All you have to do is to multiply them component by component, and then sum them up.

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = u_1 \cdot v_1 + u_2 \cdot v_2 + u_3 \cdot v_3$$

Let us see 2 concrete examples:

$$\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \vec{v} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$$



$$\vec{u} \cdot \vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} = 1 \cdot (-1) + 2 \cdot 0 + 3 \cdot 2 = 5$$

This example is a discrete case in $3D$. For 2 n-dimensional vectors (still in the discrete case), the dot product is defined as:

$$\vec{u} \cdot \vec{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 \cdot v_1 + u_2 \cdot v_2 + \ldots + u_n \cdot v_n =$$

$$= \sum_{i=1}^{n} u_i \cdot v_i$$

As we saw before, vectors are not only arrows, they can be functions as well. Consider 2 real-valued functions $p(x)$ and $q(x)$. We are not in the discrete case anymore. This is a *continuous case*, since $p(x)$ and $q(x)$ are continuous functions. They can be thought of as vectors having infinitely many "components" indexed by x instead of i. The index x is a real number, and that is why it is a sort of continuous vector with infinite components. Meanwhile, the previous example had an index i, as an integer, and that is why it was discrete.

When we have an infinite sum over a continuous index, the sum operator in the discrete case becomes the integral symbol in this continuous case $\left(\Sigma \rightarrow \int \right)$.

$$< p(x) \, ; q(x) > = \int_a^b p(x) \, q(x) \, dx$$

dot product notation
in continuous cases

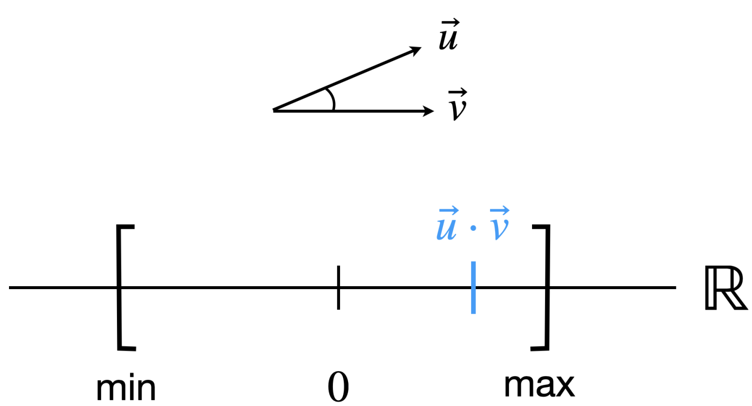
Of course, in order to fully understand the meaning of this integral one must know Calculus, but we won't discuss it in detail here. Just remember how powerful, and flexible at the same time, the dot product is.

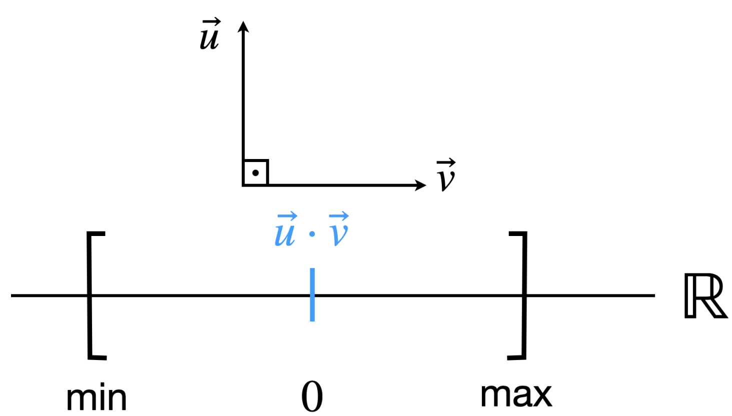
Ok, but why do we do that? I mean, why is the dot product so important after all? Well, there are many interpretations of the dot product, but one of them is that it provides a way of measuring how much 2 vectors are <u>not</u> orthogonal to each other. Let us see what this is supposed to mean.

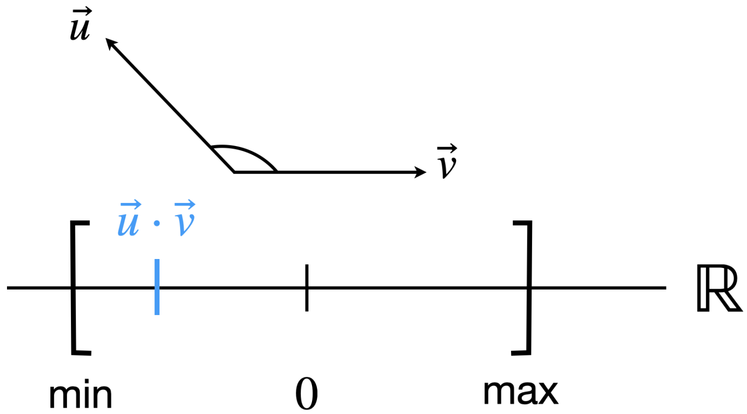
If there are 2 vectors $\vec{u}$ and $\vec{v}$ in the real plane, then we say that they are orthogonal to each other if their dot product is *zero*: $\vec{u} \cdot \vec{v} = 0$.

$\vec{u} \cdot \vec{v} \neq 0 \Rightarrow \vec{u}$ *and* $\vec{v}$ are not orthogonal to each other.

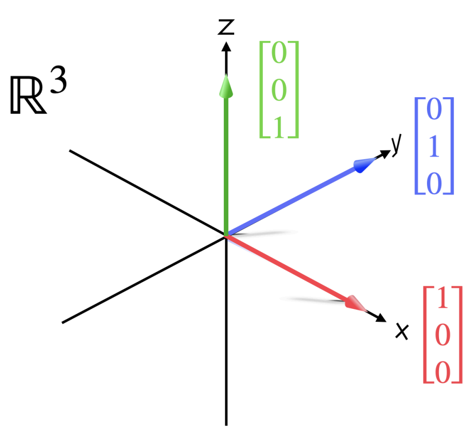
If the result of their dot product is a number that is very far from zero, we can say (loosely speaking, of course) that these vectors are "very non-orthogonal to each other":







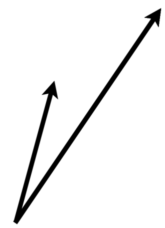
So – going back to the concept of a basis – the most common bases are the ones that are composed of orthogonal vectors, with respect to each other (and consequently are also linearly independent, of course). The classic example is the basis $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ of the Euclidean $3D$ space.



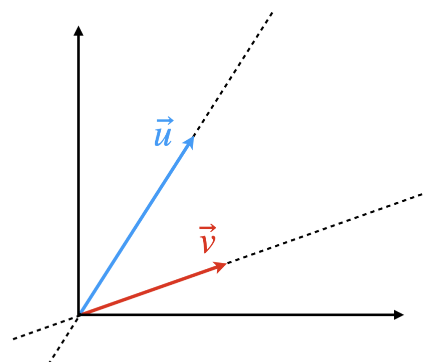
Ok, we saw a lot of things. Let us do a quick recap before moving on.

We saw the concepts of:

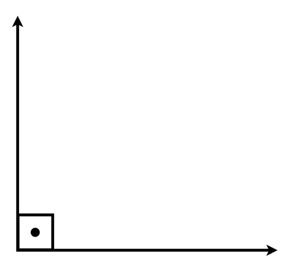
- **Vectors and their transformations**



- **Linear independence**

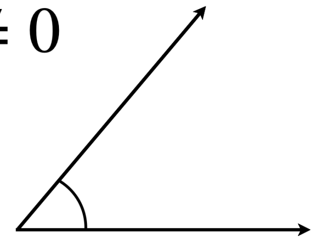


- **Orthogonality**

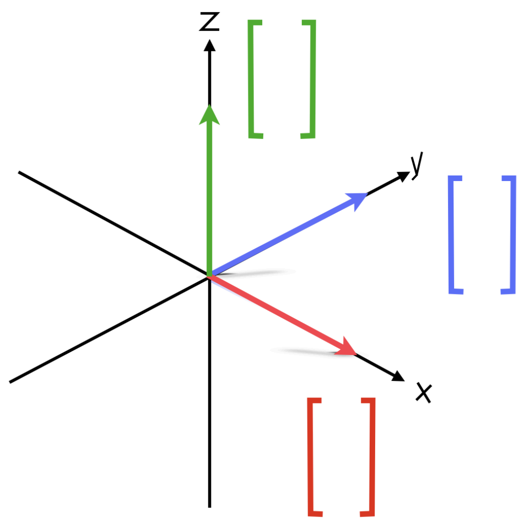


- **Dot product**

$$\vec{u} \cdot \vec{v} \neq 0$$



- **Basis vectors**



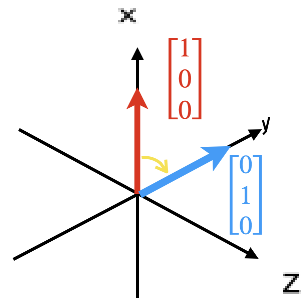
Next we will talk about some important *Matrix Transformations*. Right after that we will discuss *Systems of Linear Equations*, and we will end up with *Eigenvalues and Eigenvectors*.

<u>Matrix Transformations</u>:

(I) *Rotation*: to rotate a vector in $3D$ space (without changing its magnitude), we use rotation matrices. Just to illustrate it, let us rotate a vector $\vec{v}$ about the z-axis:

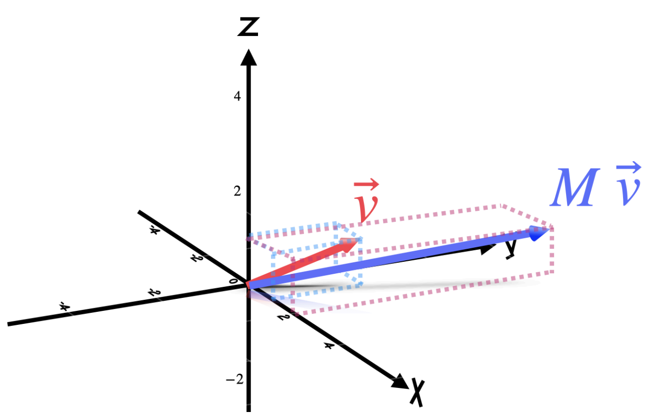
$$\vec{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad R_z\,\vec{v} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$R_z = \begin{bmatrix} \cos(90^\circ) & -\sin(90^\circ) & 0 \\ \sin(90^\circ) & \cos(90^\circ) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



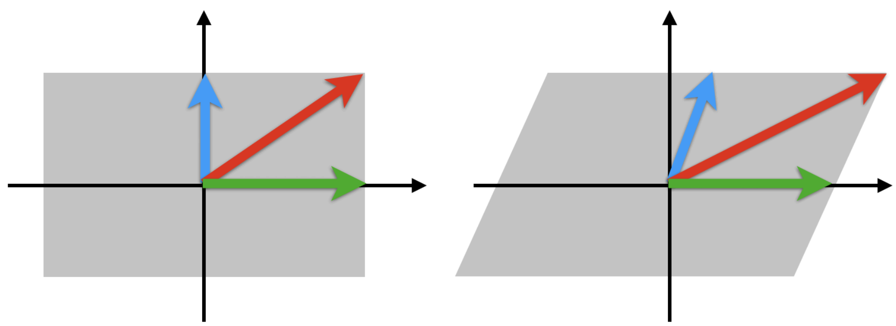
(II) *Scaling*: a scaling matrix scales a vector. But what does it mean? It means that this matrix, when acting upon a vector, changes the vector's magnitude in a particular direction. For example, the matrix $M = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ scales the vector $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$:

$$M \cdot \vec{v} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix}$$



Thus, the vector is scaled by 2 in the x-direction and by 3 in the y-direction.

(III) *Shearing*: This is a linear transformation that shifts one part of an object more than another, while preserving parallelism.



Matrices are very important in Linear Algebra, especially when solving systems of linear equations, because they give us a compact way to manipulate these systems.

But, the question is: What are *Systems of Linear Equations*?

It is basically a bunch of equations involving the same set of variables, such that all of them must be satisfied simultaneously in order to have a valid solution. Also, these equations are special, they are linear, which means that no variables are multiplied together, or raised to powers, or appears inside nonlinear functions, like $sine$, $cosine$ or *logarithms*.

Let us see a concrete example:

$$\begin{cases} 2x + y - z = 5 \\ -x + 3y + 2z = 4 \\ 3x - y + 4z = 7 \end{cases} \begin{matrix} \implies 2x + y - 5 = z \\ \implies -4 + 3y + 2z = x \\ \implies 3x - 7 + 4z = y \end{matrix} \implies$$

$$\implies x = 3y + 2 \cdot (2x + y - 5) - 4 \implies$$

$$\implies \boxed{x = -\frac{5}{3}y + \frac{14}{3}} \rightarrow y = 3x + 4z - 7 \implies$$

$$\implies y = 3 \cdot \left(-\frac{5}{3}y + \frac{14}{3}\right) + 4z - 7 \implies \boxed{y = \frac{2}{3}z + \frac{7}{6}}$$

$$z = 2x + y - 5 \implies z = 2 \cdot \left[-\frac{5}{3} \cdot \left(\frac{2}{3}z + \frac{7}{6}\right) + \frac{14}{3}\right] + \left(\frac{2}{3}z + \frac{7}{6}\right) - 5 \implies \boxed{z = \frac{29}{46}}$$

$$x = -\frac{5}{3}y + \frac{14}{3} \implies \boxed{x = \frac{93}{46}} \qquad y = \frac{2}{3}z + \frac{7}{6} \implies \boxed{y = \frac{73}{46}}$$

Isolating variables and substituting them into other equations in the system allows us to find solutions, of course as long as it is possible.

Ok, but how do these systems have anything to do with what we've seen so far? I mean, what is the connection between it and vectors and matrices?

Well, the same system can be represented in matrix form:

$$\begin{cases} 2x + y - z = 5 \\ -x + 3y + 2z = 4 \\ 3x - y + 4z = 7 \end{cases} \iff A\,\vec{x} = \vec{b}$$

$$\begin{bmatrix} 2 & 1 & -1 \\ -1 & 3 & 2 \\ 3 & -1 & 4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 7 \end{bmatrix}$$

By "translating" the system in matrix form, we can apply many systematic methods to find solutions, like *Gauss elimination*, *matrix inverses*, and *computational algorithms*.

Let us quickly see what *Gauss elimination* is.

Gaussian elimination is a method to solve a system of linear equations by transforming the system's *augmented matrix* into *row echelon form – i.e. a simplified version of a matrix where each leading entry (the first nonzero numbers in each row) is to the right of the leading entry in the row above it, and rows with only zeros are at the bottom.* These are the steps to achieve that:

1. **Write the augmented matrix**: This is a matrix that combines the coefficients and constants of a system of linear equations into a single matrix for solving it.
2. **Row reduction**:
   - Use elementary row operations (swap rows, multiply a row by a scalar, or add/subtract rows) to form zeros below the *pivot positions – i.e. the location in a matrix that corresponds to the first nonzero entry in a row after performing row reductions.*

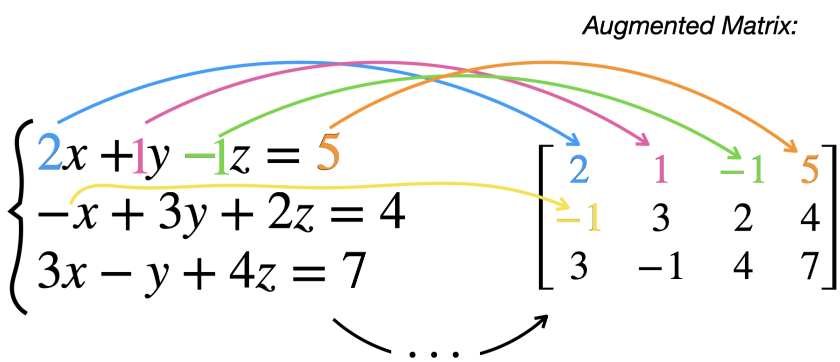
3. **Back-substitution**: Solve for the variables starting from the bottom row of the matrix.

Ok, this is pretty abstract, I know. And that's why we will see a concrete example now. But, anyway, I invite you to keep on coming back to this sort of "recipe" so that you better understand the algorithm.

Example: We want to solve the following system of linear equations:

$$\begin{cases} 2x + y - z = 5 \\ -x + 3y + 2z = 4 \\ 3x - y + 4z = 7 \end{cases}$$

1. **Write the augmented matrix**:



2. **Row reduction**:

$$\begin{matrix} R_1: \\ R_2: \\ R_3: \end{matrix}\begin{bmatrix} 2 & 1 & -1 & 5 \\ -1 & 3 & 2 & 4 \\ 3 & -1 & 4 & 7 \end{bmatrix} \xRightarrow[(R_2 \to R_1 + 2\cdot R_2)]{} \begin{matrix} R_1: \\ R_2: \\ R_3: \end{matrix}\begin{bmatrix} 2 & 1 & -1 & 5 \\ 0 & 7 & 3 & 13 \\ 3 & -1 & 4 & 7 \end{bmatrix}$$

$$\xRightarrow[\left(R_3 \to \frac{1}{3}\cdot R_3\right)]{} \begin{matrix} R_1: \\ R_2: \\ R_3: \end{matrix}\begin{bmatrix} 2 & 1 & -1 & 5 \\ 0 & 7 & 3 & 13 \\ 1 & -\frac{1}{3} & \frac{4}{3} & \frac{7}{3} \end{bmatrix} \xRightarrow[(R_3 \to R_1 - 2\cdot R_3)]{}$$

$$\Rightarrow \begin{matrix} R_1: \\ R_2: \\ R_3: \end{matrix}\begin{bmatrix} 2 & 1 & -1 & 5 \\ 0 & 7 & 3 & 13 \\ 0 & \frac{5}{3} & -\frac{11}{3} & \frac{1}{3} \end{bmatrix} \xRightarrow[\left(R_2 \to \frac{1}{7}\cdot R_2\right)]{\left(R_3 \to \frac{3}{5}\cdot R_3\right)} \begin{matrix} R_1: \\ R_2: \\ R_3: \end{matrix}\begin{bmatrix} 2 & 1 & -1 & 5 \\ 0 & 1 & \frac{3}{7} & \frac{13}{7} \\ 0 & 1 & -\frac{11}{5} & \frac{1}{5} \end{bmatrix}$$

$$\xRightarrow[\left(R_1 \to \frac{1}{2}\cdot R_1\right)]{(R_3 \to R_2 - R_3)} \begin{matrix} R_1: \\ R_2: \\ R_3: \end{matrix}\begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & \frac{5}{2} \\ 0 & 1 & \frac{3}{7} & \frac{13}{7} \\ 0 & 0 & \frac{92}{35} & \frac{58}{35} \end{bmatrix} \xRightarrow[\left(R_3 \to \frac{35}{92}\cdot R_3\right)]{}$$

$$\Rightarrow \begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & \frac{5}{2} \\ 0 & 1 & \frac{3}{7} & \frac{13}{7} \\ 0 & 0 & 1 & \frac{58}{92} \end{bmatrix} \Longrightarrow \begin{cases} x + \frac{y}{2} - \frac{z}{2} = \frac{5}{2} & \Longrightarrow \boxed{x = \frac{93}{46}} \\ y + \frac{3}{7}z = \frac{13}{7} & \Longrightarrow \boxed{y = \frac{73}{46}} \\ z = \frac{58}{92} & \Longrightarrow \boxed{z = \frac{29}{46}} \end{cases}$$

leading entries

The best way to appreciate how powerful the matrix formalism is, is in the context of large systems (with 10, 100, or even millions of equations with the same number of variables).

$$
\begin{cases}
3x_1 + 2x_2 - x_3 + 4x_4 + 5x_5 - 2x_6 + 7x_7 + x_8 - 3x_9 + 6x_{10} = 12, \\
-2x_1 + 4x_2 + 3x_3 - x_4 + 6x_5 - x_6 + 5x_7 - 2x_8 + x_9 - 4x_{10} = -7, \\
x_1 - x_2 + 2x_3 + 3x_4 - 4x_5 + 5x_6 - x_7 + 6x_8 - 2x_9 + 3x_{10} = 8, \\
4x_1 + x_2 - 3x_3 + 2x_4 + x_5 - 6x_6 + 4x_7 + x_8 + 2x_9 - x_{10} = 5, \\
x_1 + 2x_2 + x_3 - 4x_4 + 3x_5 - x_6 + 5x_7 - x_8 + 6x_9 + 4x_{10} = 10, \\
-3x_1 + x_2 + 4x_3 - 2x_4 + x_5 + 6x_6 - 3x_7 + 2x_8 - 5x_9 + x_{10} = 3, \\
2x_1 - x_2 + 5x_3 + 3x_4 - x_5 - 2x_6 + 4x_7 + x_8 - x_9 + 6x_{10} = -2, \\
x_1 + 3x_2 - 4x_3 + x_4 + 2x_5 - x_6 + 6x_7 - 2x_8 + 3x_9 - 5x_{10} = 9, \\
-2x_1 + 5x_2 + x_3 - 3x_4 + 4x_5 - 6x_6 + x_7 + 2x_8 - x_9 + x_{10} = -4, \\
3x_1 - x_2 + x_3 + 2x_4 - 5x_5 + 4x_6 - 2x_7 + 6x_8 + x_9 - 3x_{10} = 11
\end{cases}
$$

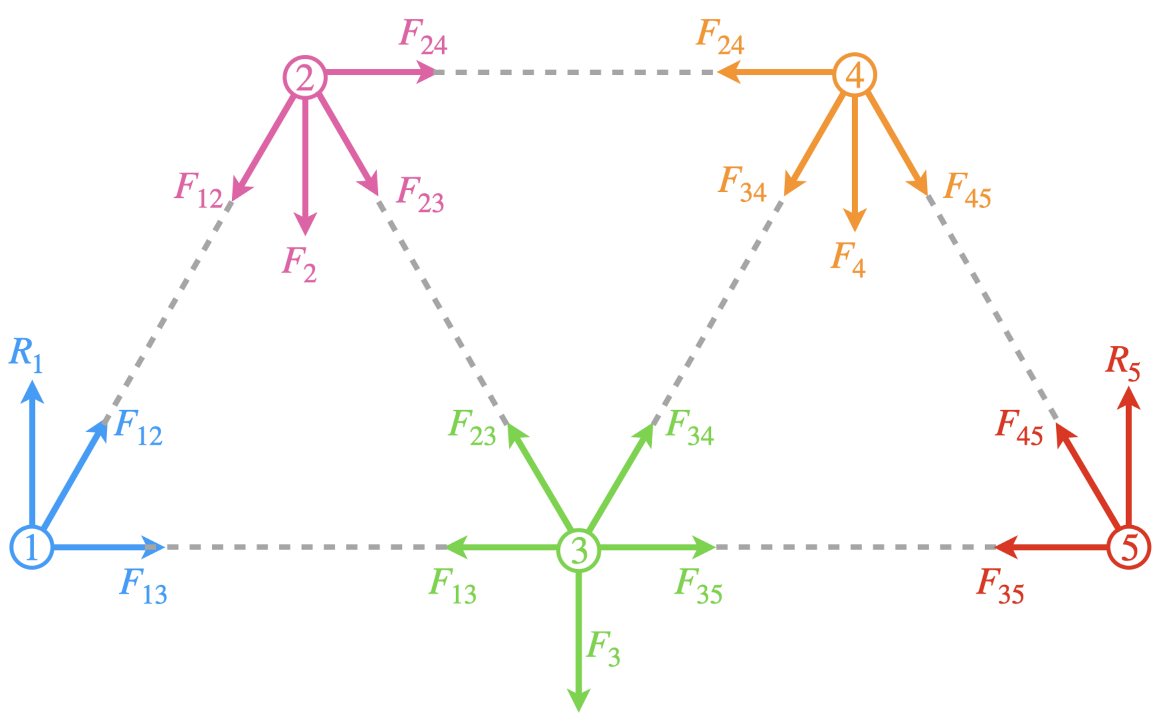
(Good luck trying to solve it – email us the result later 🤣)
(dibeos.contact@gmail.com)

Just imagine the nightmare that would be trying to solve these large systems by substitution! I mean, it is technically possible… but extremely impractical!

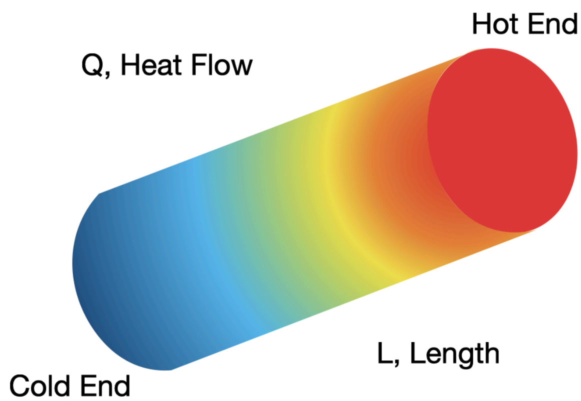
$$
\begin{bmatrix}
3 & 2 & -1 & 4 & 5 & -2 & 7 & 1 & -3 & 6 & 12 \\
-2 & 4 & 3 & -1 & 6 & -1 & 5 & -2 & 1 & -4 & -7 \\
1 & -1 & 2 & 3 & -4 & 5 & -1 & 6 & -2 & 3 & 8 \\
4 & 1 & -3 & 2 & 1 & -6 & 4 & 1 & 2 & -1 & 5 \\
1 & 2 & 1 & -4 & 3 & -1 & 5 & -1 & 6 & 4 & 10 \\
-3 & 1 & 4 & -2 & 1 & 6 & -3 & 2 & -5 & 1 & 3 \\
2 & -1 & 5 & 3 & -1 & -2 & 4 & 1 & -1 & 6 & -2 \\
1 & 3 & -4 & 1 & 2 & -1 & 6 & -2 & 3 & -5 & 9 \\
-2 & 5 & 1 & -3 & 4 & -6 & 1 & 2 & -1 & 1 & -4 \\
3 & -1 & 1 & 2 & -5 & 4 & -2 & 6 & 1 & -3 & 11
\end{bmatrix}
$$

Some practical applications are:

- In Structural Analysis (which is a branch of Engineering), where solutions of thousands of linear equations are used to model forces in bridges, skyscrapers, and aircraft structures.



- In Physics, in order to simulate phenomena involving millions of variables and equations, like heat distribution and stress analysis in materials.



- In Data Science and Machine Learning, for the obvious reason that matrices are great for storing and manipulating data. Linear regression is one of the simplest machine learning models, where we predict an output vector $\vec{y}$ using a linear combination of features in the data matrix X:

$$\vec{y} = X \cdot \vec{w}$$

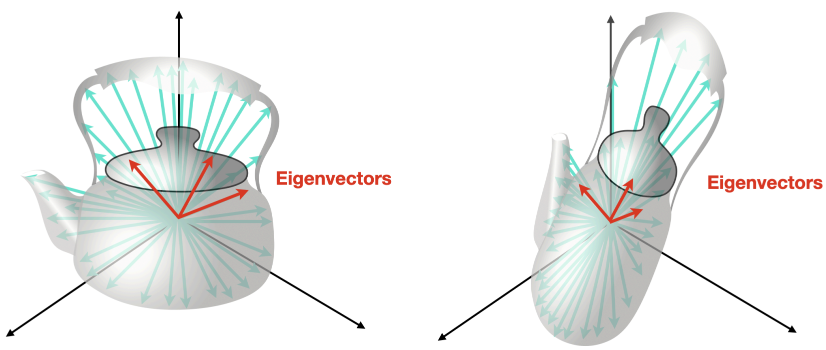
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ X_{31} & X_{32} & \dots & X_{3m} \\ \vdots & & & \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Solving for $\vec{w}$ involves minimizing the error:

$$\min_{\overrightarrow{w}} \|X\,\overrightarrow{w} - \vec{y}\|$$

One of the most useful applications of matrices and linear systems of equations is in the study of *Eigenvalues and Eigenvectors*.

Let us start with an example in order to create intuition behind the math:



Notice that all of the green vectors, which are uniquely associated to each point on the surface of the kettle, completely change their lengths and their angles with respect to the horizontal xy-plane (or yz-plane or xz-plane), after the transformation. However, out of all these vectors, there are 3 of them that are special. These are the red vectors, and they are special because after the transformation they do not change their angles, but their lengths do change. So, they are scaled by specific amounts. The real numbers that represent each of these scalings factors, for each of the red vectors, are usually denoted with the greek letter 'λ' (lambda), and they are formally called *Eigenvalues*. But, again, this is not true for every vector here, just for a specific type of vector. These special vectors (shown in red here) are called *Eigenvectors*.

Let us see this process with some numbers involved, so that we can better grasp the idea. We have a linear transformation matrix in $\mathbb{R}^3$:

$$A = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

In order to find the eigenvalues and eigenvectors of this transformation we need to solve the equation:

$A\vec{v} = \lambda I\vec{v}$, where *I* is the identity matrix $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, i.e. the matrix that (effectively) does "nothing", and so just returns the vector $\vec{v}$.

The left-hand side of this equation gives us every new vector after applying the transformation. The right-hand side imposes that this new vector must be *identical* to itself (that's why we use the identity matrix) multiplied by a scalar (λ) that measures the amount of which the length of the original vector changed with respect to the new vector after the transformation. In other words, this equation is telling us that, out of all the vectors that were transformed by the matrix, consider <u>only</u> the red ones here. In order to solve this equation, we start by looking for the eigenvalues first:

$$\boxed{A\vec{v} = \lambda I\vec{v}} \Longrightarrow A\vec{v} - \lambda I\vec{v} = \vec{0} \Longrightarrow (A - \lambda I)\vec{v} = \vec{0}$$

$$\Longrightarrow \det(A - \lambda I) = 0 \Longrightarrow \cdots \Longrightarrow$$

$$\Longrightarrow \det\begin{bmatrix} 4-\lambda & 1 & 0 \\ 1 & 4-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{bmatrix} = 0 \Longrightarrow \cdots \Longrightarrow$$

$$\Longrightarrow (3-\lambda)\cdot\left[(4-\lambda)^2 - 1\right] = 0 \Longrightarrow \begin{cases} \lambda_1 = 5 \\ \lambda_2 = 3 \\ \lambda_3 = 3 \end{cases}$$

Now that we found out what the eigenvalues of this transformation are – i.e. by "how much" the red arrows were scaled – we can calculate their associated eigenvectors:

$$A = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\boxed{A\vec{v} = \lambda I\vec{v}} \Longrightarrow (A - \lambda I)\vec{v} = \vec{0}$$

$$\xrightarrow{\lambda_1 = 5} (A - 5\cdot I)\,\vec{v}_1 = \vec{0} \Longrightarrow$$

$$\xrightarrow{\lambda_2 = 3} (A - 3\cdot I)\,\vec{v}_2 = \vec{0} \Longrightarrow$$

$$\xrightarrow{\lambda_3 = 3} (A - 3\cdot I)\,\vec{v}_3 = \vec{0} \Longrightarrow$$

$$\Longrightarrow \begin{bmatrix} 4-5 & 1 & 0 \\ 1 & 4-5 & 0 \\ 0 & 0 & 3-5 \end{bmatrix}\cdot\vec{v_1} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} 4-3 & 1 & 0 \\ 1 & 4-3 & 0 \\ 0 & 0 & 3-3 \end{bmatrix}\cdot\vec{v_2} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} 4-3 & 1 & 0 \\ 1 & 4-3 & 0 \\ 0 & 0 & 3-3 \end{bmatrix}\cdot\vec{v_3} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

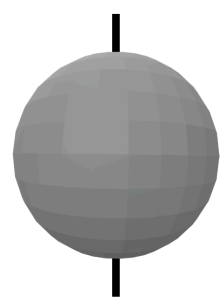
$$\longrightarrow \boxed{\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \vec{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}$$

Cool Application of Eigenvalues and Eigenvectors to Quantum Mechanics: (The Spin of an Electron)

In quantum mechanics, the *spin of an electron* can be described using matrices and eigenvalues. The spin along the z-axis is represented by the *Pauli matrix*:

$$S_z = \frac{\hbar}{2} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

, where ħ is the reduced Planck's constant.



We want to find the possible measurement outcomes (*eigenvalues*) and the states (*eigenvectors*) associated with these outcomes. In order to to that, we look for solutions of the following equation:

$$\boxed{S_z\vec{v} = \lambda\vec{v}} \implies S_z\vec{v} - \lambda\vec{v} = \vec{0} \implies \left(S_z - \lambda I\right)\vec{v} = \vec{0} \implies$$

$$\implies det\left(S_z - \lambda I\right) = 0 \implies det\begin{bmatrix} \frac{\hbar}{2} - \lambda & 0 \\ 0 & -\frac{\hbar}{2} - \lambda \end{bmatrix} = 0$$

$$\implies \left(\frac{\hbar}{2} - \lambda\right)\left(-\frac{\hbar}{2} - \lambda\right) = 0 \implies -\frac{\hbar^2}{4} + \lambda^2 = 0$$

$$\implies \lambda^2 = \frac{\hbar^2}{4} \implies \boxed{\lambda_1 = \frac{\hbar}{2}} \wedge \boxed{\lambda_2 = -\frac{\hbar}{2}}$$

Eigenvalues!

$$\boxed{\lambda_1 = \frac{\hbar}{2}} \implies S_z\vec{v} = \frac{\hbar}{2}\vec{v} \implies \left(S_z - \frac{\hbar}{2}I\right)\vec{v} = \vec{0} \implies$$

$$\implies \begin{bmatrix} \frac{\hbar}{2} - \frac{\hbar}{2} \; (0) & 0 \\ 0 & -\frac{\hbar}{2} - \frac{\hbar}{2} \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies$$

$$\implies \begin{bmatrix} 0 \\ -2\frac{\hbar}{2}v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{cases} 0 = 0 \implies v_1 = \text{any scalar} \\ -\frac{\hbar}{2}v_2 = 0 \implies v_2 = 0 \end{cases}$$

(to simplify $v_1 = 1$)

$$\therefore \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\boxed{\lambda_2 = -\frac{\hbar}{2}} \Longrightarrow S_z\vec{v} = -\frac{\hbar}{2}\vec{v} \Longrightarrow \left(S_z + \frac{\hbar}{2}I\right)\vec{v} = \vec{0} \Longrightarrow$$

$$\Longrightarrow \begin{bmatrix} \frac{\hbar}{2} + \frac{\hbar}{2} & 0 \\ 0 & \cancelto{0}{-\frac{\hbar}{2} + \frac{\hbar}{2}} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longrightarrow$$

$$\Longrightarrow \begin{bmatrix} \hbar v_1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longrightarrow \begin{cases} \hbar v_1 = 0 \Longrightarrow v_1 = 0 \\ 0 = 0 \Longrightarrow v_2 = \text{any scalar} \\ \quad \text{(to simplify } v_2 = 1\text{)} \end{cases}$$

$$\therefore \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Therefore, the 2 eigenvectors are:

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \wedge \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Please, if you find this document useful, let us know. Or if you found typos and things to improve, let us know as well. Your feedback is very important to us. We're working hard to deliver the best material possible. Contact us at: dibeos.contact@gmail.com